

The molecular architecture of the TNF superfamily

Jean-Luc Bodmer, Pascal Schneider and Jürg Tschopp

Ligands of the TNF (tumour necrosis factor) superfamily have pivotal roles in the organization and function of the immune system, and are implicated in the aetiology of several acquired and genetic diseases. TNF ligands share a common structural motif, the TNF homology domain (THD), which binds to cysteine-rich domains (CRDs) of TNF receptors. CRDs are composed of structural modules, whose variation in number and type confers heterogeneity upon the family. Protein folds reminiscent of the THD and CRD are also found in other protein families, raising the possibility that the mode of interaction between TNF and TNF receptors might be conserved in other contexts.

Metazoan organisms consist of an intricate and ordered society of individual cells that must communicate to maintain and regulate their functions. This is achieved through a complex and highly regulated network of hormones, chemical mediators, chemokines and other cytokines, acting as ligands for intra- or extracellular receptors. Ligands and receptors of the tumour necrosis factor (TNF) superfamily are examples of signal transducers whose integrated actions impinge principally on the development, homeostasis and adaptive responses of the immune system. Despite their varied and pleiotropic actions, members of the TNF ligand and receptor (TNFR) families have remarkably similar structures, and their mode of interaction is conserved. The aim of this review is to provide an overview of the molecular architecture and the modular organization of the TNF and TNFR gene superfamilies.

The TNF family

The TNF ligand family comprises 18 genes encoding 19 type II (i.e. intracellular N terminus and extracellular C terminus) transmembrane proteins, characterized by a conserved C-terminal domain coined the 'TNF homology domain' (THD) (Fig. 1). This trimeric domain is responsible for receptor binding and its sequence identity between family members is ~20–30%. Although most ligands are synthesized as membrane-bound proteins, soluble forms can be generated by limited proteolysis (Fig. 1). Distinct proteases are involved in this process, depending on the ligand: metalloproteases of the ADAM (a disintegrin and metalloproteinase domain) family act on TNF and RANKL ligands [1,2], matrixlysin acts on Fas ligand (FasL) [3], and members of the subtilisin-like furin family act on BAFF, EDA, TWEAK and APRIL—members of the TNF family [4,5]. Solubilization is essential for the physiological function of some ligands; mutation in

the furin recognition sequence of EDA is a frequent cause of the genetic disorder X-linked hypohidrotic ectodermal dysplasia (XLHED) [4,6]. By contrast, the shedding of some ligands inhibits their function. For instance, the cytotoxic activity of FasL is dramatically downregulated upon cleavage [7]. The N terminus of lymphotoxin α (LT α) resembles a signal peptide, making its conversion to a soluble form extremely efficient. Consequently, LT α is never found at the cell surface except when it is associated with membrane-bound LT β as LT $\alpha_1\beta_2$ heterocomplexes [8] (Fig. 1). Processing of TNF-related apoptosis-inducing ligand (TRAIL) by a cysteine protease has been proposed [9], but the resulting soluble form seems to be too small to retain a functional THD.

Ligands of the TNF family control and orchestrate the immune and inflammatory responses at several levels (recently reviewed in Ref. [10]). During development, TNF ligands such as TNF, LT α , LT β and RANKL provide crucial signals for the morphogenesis of secondary lymphoid organs [10,11]. In addition, the grooming and proper activation of immune precursor cells to fully competent effectors is dependent on several other TNF family members such as BAFF and CD40L for B lymphocytes [12–14]; 4-1BBL, OX40L and CD27L for T lymphocytes [15]; and CD40L and RANKL for dendritic cells [16,17]. Pro-apoptotic members of the family (e.g. TNF, FasL and TRAIL) contribute to the function of cytotoxic effector cells and participate in the homeostasis of the lymphoid compartment by evoking activation-induced cell death in immune effector cells that have fulfilled their function [18]. Recent evidence indicates that other TNF family ligands regulate the development and differentiation of epithelial structures (the EDA ligand), endothelial cells (VEG1 and TWEAK) and bone-resorbing osteoclasts (RANKL and TNF) [10].

TNF family ligands and receptors are associated with several disease conditions that result from acquired processes or genetic defects. Acquired acute or chronic inflammatory conditions such as septic shock or rheumatoid arthritis result from excessive or inappropriate TNF expression [19]. Mutations in TNF ligands and/or receptors have been described in five hereditary diseases: hyper IgM syndrome (HIM, CD40L), type I autoimmune lymphoproliferative syndrome (ALPS, Fas/FasL), TNF-R1-associated periodic fever syndrome (TRAPS, TNF-R1),

Jürg Tschopp*
Pascal Schneider
Jean-Luc Bodmer
Institute of Biochemistry,
University of Lausanne,
Chemin des
Boveresses 155,
1066 Epalinges,
Switzerland.
*e-mail: Jürg.Tschopp@
ib.unil.ch

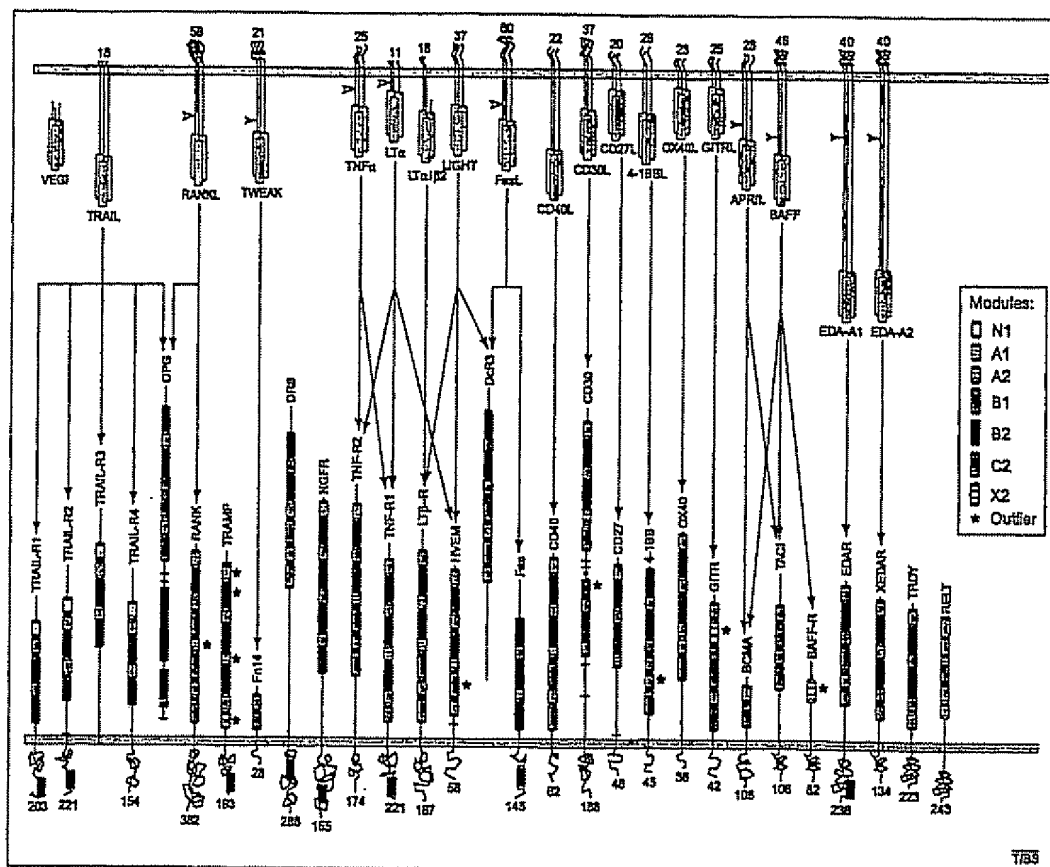


Fig. 1. Interactions between ligands and receptors of the human tumour necrosis factor (TNF) family. TNF ligands (top) and TNF receptors (bottom). The TNF ligands are represented as type II homo- or heterotrimeric transmembrane proteins (with the exception of VEGF, which lacks a predicted transmembrane domain and is therefore drawn as a soluble ligand). TNF homology domains (THDs) are shown as green boxes, filled black arrowheads indicate processing by furin family members, and open black arrowheads by other types of proteases. The TNF receptors are typically type I or type III transmembrane proteins, but also occur as glycosylated or soluble proteins. N1, A1, A2, B1, B2, C2 and X2 modules are colour-coded as shown in the inset. The positions of individual cysteines are indicated by horizontal bars, and stars show modules whose cysteine pattern does not conform entirely to that of canonical A, B, C and N modules. The lengths of intracellular domains are indicated for each ligand and each receptor, and the intracellular homology domains, known as the 'death domains', are indicated as red boxes. Red arrows show documented interactions. An interaction between TWEAK and TRAMP has been reported [53] but has not yet been confirmed. Some of the ligands and receptors have several commonly used names: FasL/CD95, TRAIL/Apo-2L, RANKL/OPG/TRANCE, BAFF/BLyS/TALL-1, Fas/CD95, TRAIL-R1/DR4, TRAIL-R2/DR5, TRAIL-R3/DR6, TRAIL-R4/DR2, TRAMP/DR3 and TRCY/DAI. For the official TNF superfamily (TNFSF) nomenclature and additional synonyms, consult <http://www.genenames.org/nomenclature/genefamily/tnf.htm>

hypohidrotic ectodermal dysplasia (HED, EDA/EDAR) and familial expansile osteolysis (FEO, RANK) [10]. It is likely that other links between TNF members and diseases will be uncovered in the future.

Structural features of TNF family ligands

The THD is a 150 amino acid long sequence containing a conserved framework of aromatic and hydrophobic residues (Fig. 2). To date, atomic-level

structures are available for the THD of TNF [20,21], LT α [22], CD40L [23] and TRAIL [24–27]. THDs share a virtually identical tertiary fold and associate to form trimeric proteins (Fig. 3a). The THDs are β -sandwich structures containing two stacked β -pleated sheets each formed by five anti-parallel β strands that adopt a classical 'jelly-roll' topology. The inner sheet (strands A, A', H, C and F) is involved in trimer contacts, and the outer sheet (strands B, B', D, E and G) is exposed at the surface. Trimeric THDs are ~ 60 Å in height and resemble bell-shaped, truncated pyramids with variable loops protruding out of a compact core of conserved anti-parallel β strands (Figs 2,3a). TRAIL is unique with respect to the AA' loop, which contains a 15 residue-long insertion that spans the whole outer surface of the monomer [24,26,27]. The trimer is assembled such that one edge of each subunit (strands E and F) is packed against the inner sheet of its neighbour, forming large and mostly hydrophobic interfaces, resulting in a very stable interaction [20,26,27]. TNF and CD40L contain a single disulfide bridge linking the CD and EF loops [20,23] (Fig. 3a). Similar disulfide links are

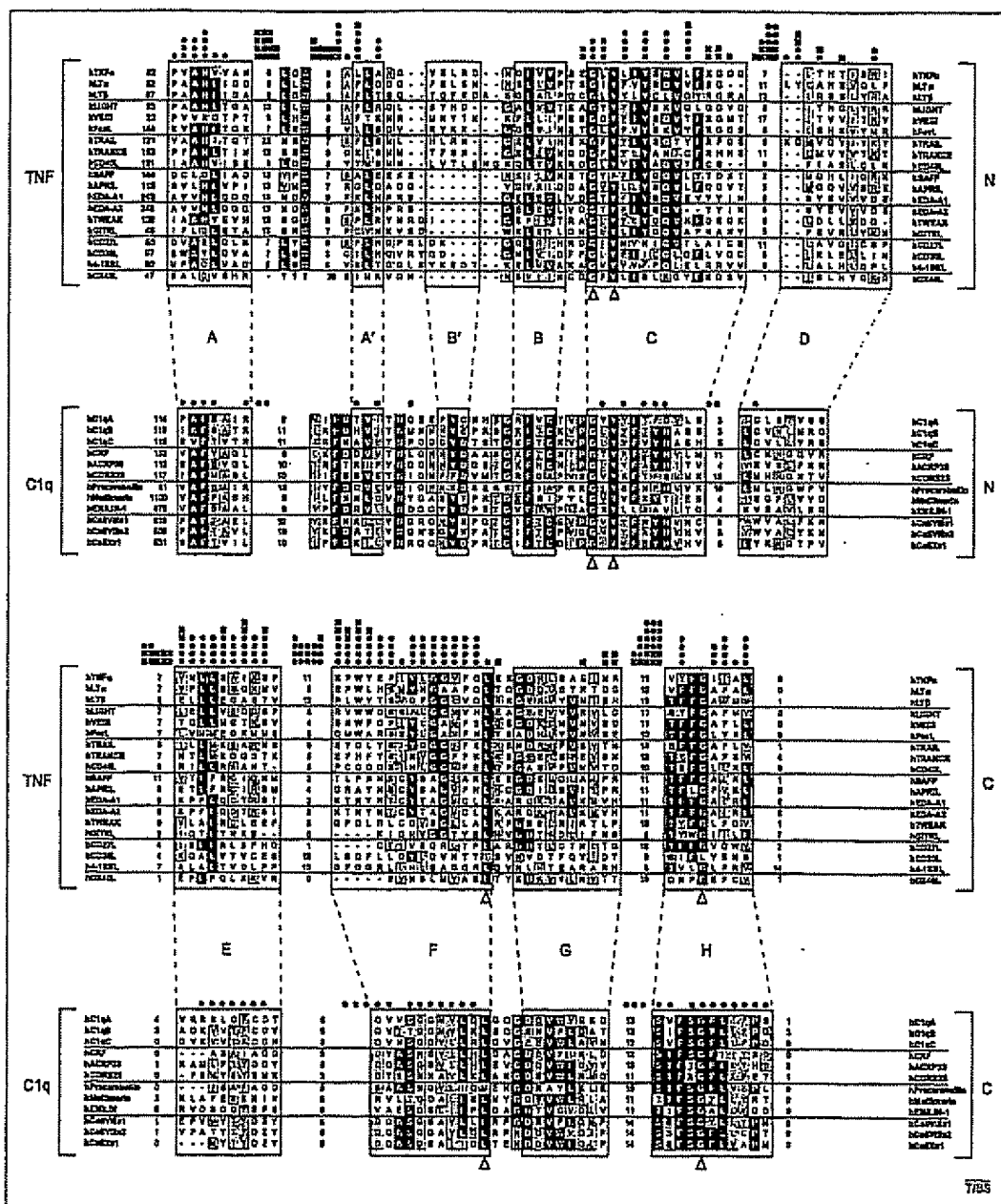


Fig. 2. Sequence alignment of the human tumour necrosis factor (TNF) and C1q superfamilies. Primary sequence alignments of the TNF homology domains (THDs) of 19 TNF ligands (including the distantly related member OX40L) and of 12 published C1q-related proteins. The alignment has been reduced to regions of significant sequence homology. Intervening loops have been omitted except for the conserved LVSW motif in the AA' loop of the TNF ligands, but their length is indicated. The individual β strands (A-H) are highlighted with boxes coloured with respect to their succession in the primary structure from red to magenta. Blue dots above the alignment indicate residues involved in monomer-monomer interface formation and

their numbers represent the frequency at which each position is found to interact in the five structures available (e.g. TNF, LT α , CD40L, TRAIL and mACRP30). Red squares represent residues involved in receptor binding in the two complex structures available (LT α -TNF-R1 and TRAIL-DR5). Arrowheads underneath the sequences point to the four conserved residues in the TNF and C1q families. The multiple sequence alignment was generated with the amino acid sequence of THD and gC1q domains using ClustalW, and was edited manually to account for structural knowledge. Identical amino acids (white text on black background) and 50% similar amino acids (on grey background) were shaded using Boxshade.

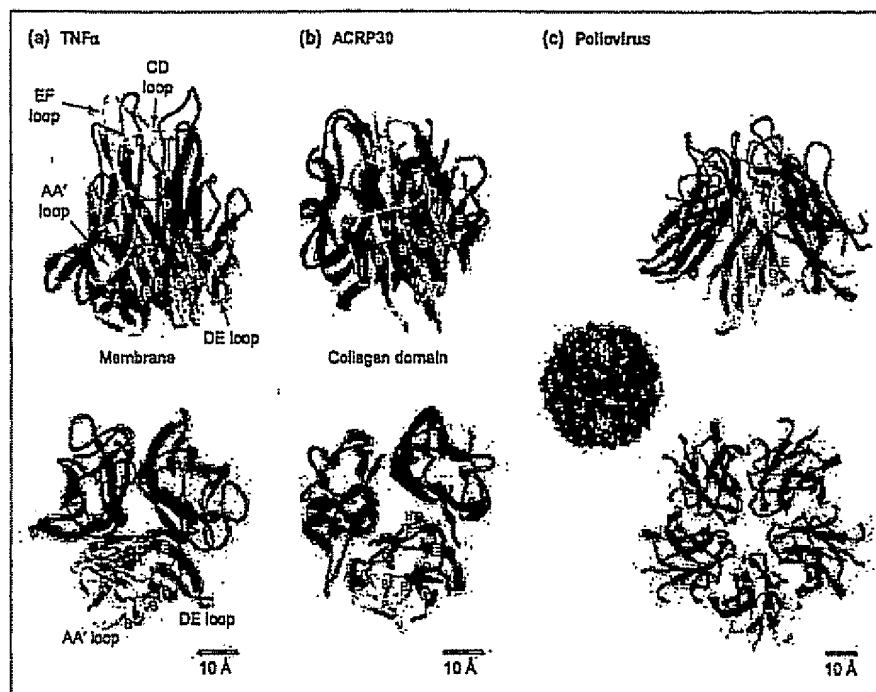


Fig. 3. The tumour necrosis factor (TNF) homology domain (THD). (a) Ribbon diagrams of the THD of human TNF seen from the side (top) and top (bottom) orientation. One monomer is highlighted and the other two are shaded. The ten anti-parallel β strands (designated A, A', B', B', C, D, E, F, G and H according to Ref. [20]) are coloured using the same code as in Fig. 2. Intervening loops are shown in white. The close proximity of N and C termini. Models are based on the PDB atomic coordinates file 1TNF [20]. (b) Structure of the gC1q domain of murine ACRP30 seen from the side (top) and top (bottom) orientation showing its similarity to the THD. Models are based on the PDB atomic coordinates file 1C28 [31]. (c) Representation of the pentameric 'jelly-roll' domain of the VP1 capsid protein (boxed in the representation of the viral particle in the centre) of the Mahoney strain of type 1 human poliovirus seen from the side (top) and top (bottom) orientation. Strands are coloured and numbered as in (a) and (b). The topological organization of the eight strands is identical to that of the THD, with the exception of the two interruptions in strands A and B. Models are based on the PDB atomic coordinates file 2PLV [35].

predicted to occur in FasL, LIGHT, VEGI, CD30L and CD27L, whereas TWEAK, EDA, APRIL and BAFF have a predicted disulfide bridge between β strands E and F. In TRAIL, a single cysteine residue (Cys230) in the EF loop is involved in the coordination of a Zn(II) ion, with each monomer contributing to one coordination position; the fourth coordination position is occupied by an internal solvent molecule or a chloride counter-ion [24–26]. This metal-binding site is unique so far in the TNF family, and affects the stability and bioactivity of TRAIL [26,28,29]. Incomplete Zn coordination, and formation of partially oxidized, disulfide-linked species of TRAIL, have recently been suggested to account for its hepatotoxicity [30].

<http://tbs.trends.com>

TNF-related structures - the C1q family

Crystallographic studies revealed that TNF and the globular gC1q domain of mouse ACRP30 have a closely related tertiary structure and trimeric organization, suggestive of an evolutionary link between the TNF and C1q families [31] (Fig. 3a,b). The human C1q gene family comprises, so far, 13 members (Fig. 2), which are characterized by the presence of a trimeric globular C-terminal domain, known as gC1q. The prototypical member of the family is C1q, a bouquet-like molecule comprising 18 chains (six each of C1qA, C1qB and C1qC) that associate into six heterotrimeric gC1q domains held together by a bundle of collagen domains. C1q recognizes immune complexes and triggers the classical complement pathway (recently reviewed in Ref. [32]). The C1q family also contains several collagenous members (CRF, ACRP30, CORS26, EMILIN-1 and -2, and collagens VII and X) and two non-collagenous members (Precerebellin and Multimerin) (Fig. 2). Many of these proteins are components of the extracellular matrix in diverse organs [32]. ACRP30 is an abundant serum protein that is synthesized by adipose tissues in response to insulin, and is downregulated in obese mouse and humans [33,34]. The homologues of ACRP30 are drastically downregulated in the serum of hibernating Siberian chipmunks, pointing to a role in energy metabolism. Indeed, ACRP30 induces weight loss in mice via activation of fatty acid catabolism [35]. Conserved residues of gC1q domains are located

within the core β strands, as previously discussed for THD domains. Although the sequence identity between the two families is reduced to only four amino acid residues (indicated with arrowheads in Fig. 2), the overall hydrophobic character of the internal β -pleated sheet is maintained in both families.

To date, there is no evidence that the mode of interaction described in the TNF family also applies to proteins containing globular C1q domains. Several C1q receptors have been described, but none of them, with the notable exception of the immune complex, binds to the globular domain. Nevertheless, the recent demonstration that the gC1q domain of ACRP30 is biologically active implies the existence of ACRP30 receptor(s), which could be structurally related to TNF receptors.

TNF-related structures - viral capsid proteins

As first noticed by Jones and Eck in 1989, the overall fold and topology of the THD is very similar to that of the capsid proteins of small spherical plant viruses (e.g. Tomato Bushy Stunt Virus and Satellite Tobacco Necrosis Virus) and mammalian picornaviruses (including the common human Rhinoviruses, the Foot-and-Mouth Disease Virus and Poliovirus), despite there being no detectable similarity at the primary sequence level [20,21]. Although these capsid proteins associate along a fivefold instead of a threefold axis of symmetry, the connectivity of their β strands is identical to that of the THD, with the exception that strands A and B are not interrupted by loops [36] (Fig. 3c). These fivefold structures appear on the virus surface as 12 broad, star-like protuberances (Fig. 3c). Although the structural relationship existing between these apparently unrelated protein families highlights the propensity of 'jelly-roll' motifs to oligomerize, there appears to be no functional conservation between TNF family members and icosahedral viral capsid proteins. Indeed, the receptors allowing entry of this class of viruses into cells do not belong to the TNF receptor family and do not bind directly to the oligomeric 'jelly-roll' structure [37].

The TNF receptor family

In humans, 29 TNF receptors have so far been identified (Fig. 1). These are primarily type I (extracellular N terminus, intracellular C terminus) transmembrane proteins, but there are exceptions to this rule: BCMA, TACI, BAFFR and XEDAR are type III transmembrane proteins (lacking a signal peptide), TRAIL-R3 is anchored by a covalently linked C-terminal glycolipid, and OPG and DcR3 lack a membrane-interacting domain and are therefore secreted as soluble proteins. Soluble receptors can also be generated by proteolytic processing (CD27, CD30, CD40, TNF-R1 and TNF-R2) [38], or by alternative splicing of the exon encoding the

transmembrane domain (Fas and 4-1BB) [39]. The essential role of these soluble receptors in modulating the activity of their cognate ligands has been well-documented (for OPG and TNF-R1 examples, see Refs [40,41]). In addition, several viral open reading frames encode receptor homologues that interact with TNF and that are believed to interfere with the onset of inflammatory responses: SVF-T2 in Shope-fibroma virus, Va53R in Vaccinia, and cytokine response modifiers CrmB, CrmC and CrmD in orthopoxviruses (reviewed in Ref. [39]). The TNF receptor family member NGFR is unique in that it binds low-affinity ligands that do not belong to the TNF family. These ligands (NGF, BDNF and neurotrophins) also engage a family of high-affinity tyrosine receptor kinases (trkA, B and C), which are unrelated to TNF receptors [42]. The existence of a *bona fide* TNF ligand for NGFR cannot be excluded at present.

Structural features of the TNF receptor family

The extracellular domains of TNF receptors are characterized by the presence of cysteine-rich domains (CRDs), which are pseudo-repeats typically containing six cysteine residues engaged in the formation of three disulfide bonds. The number of CRDs in a given receptor varies from one to four, except in the case of CD30 where the three CRDs have been partially duplicated in the human but not in the mouse sequence. The repeated and regular arrangement of CRDs confers an elongated shape upon the receptors, which is stabilized by a slightly twisted ladder of disulfide bridges (Figs 1,4). Sequence alignment of TNF receptor family members in the absence of structural information is difficult because the spacing of cysteine residues is not always conserved between receptors. Naismith and Sprang have introduced a classification based on distinct structural modules that greatly facilitates sequence comparison between TNF receptors [43]. Each module type is designated by a letter (A, B, C and N for crystallized modules, and X for modules of unknown structure), and by a numeral indicating the number of disulfide bridges it contains. A typical CRD is usually composed of an A1-B2 or A2-B1 module or, less frequently, a different pair of modules. A1 modules are 12-27 amino acids long, consist of three short β strands linked by turns, and contain a single disulfide bridge connecting strands 1 and 3, yielding a characteristic C-shaped structure (Fig. 4). A2 modules contain a second disulfide bridge linking the second and third strands without affecting the overall structure. B modules are 21-24 amino acids long and comprise three anti-parallel strands adopting an S-shaped fold reminiscent of a paper clip (Fig. 4). In this case, the fold is constrained by two entangled disulfide bridges linking strands 1 and 3 in B2 modules. The first disulfide bridge is replaced by a hydrogen bond in B1 modules [43].

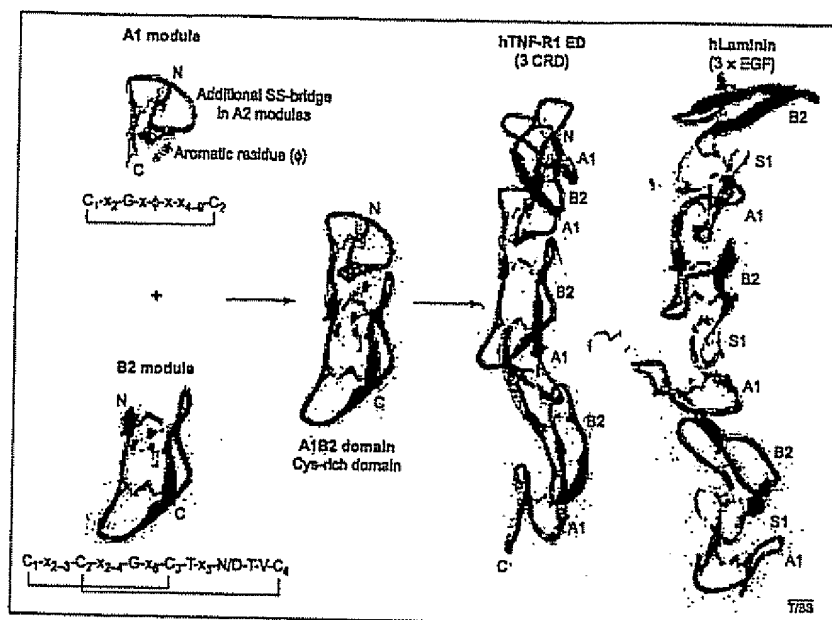


Fig. 4. Modular organization of the extracellular domain of the tumour necrosis factor (TNF) receptors. Ribbon representation of an A1 module (light blue, top left) and a B2 module (dark blue, bottom left) along with their consensus sequences. In the middle, the two modules are combined to generate a cysteine rich domain (CRD). The full extracellular domain contains several CRDs stacked on top of each other. The tertiary structure of the epidermal growth factor (EGF) domains of human laminin- γ 1 chain is remarkably similar, except that an additional loop is inserted between CRDs (S1, shown in white). Models are based on the PDB atomic coordinate files 1TNF (TNF receptor) [22] and 1KLD (laminin γ) [48].

The structure of A and B modules is also reflected at the level of the primary sequence by the conservation of a few non-cysteine residues. Other modules are less frequent. So far, the N-terminal N1 modules have been found only in the TRAIL receptors, in which they precede the first A1-B2 CRD. Structurally, the N1 module resembles the second half of a B module [24,25,27]. The fourth CRD of TNF-R1 contains an A1-C2 module pair, in which the cysteine connectivity of C2 is distinct from that of a B2 module. TACI, BCMA and Fn14 also contain putative A1-C2 CRDs, but these remain to be demonstrated at the structural level. Finally, we have collectively designated, as X2, four unrelated modules of unknown structure that are found in TRAMP, GITR, BAFFR and viral CrmC. The recently described BAFF receptor (BAFFR) [44] contains a single X2 module whose sequence resembles an A module entangled with the beginning of a B module. More structural work is needed to understand the molecular interaction of this receptor with BAFF. TNF receptors are often viewed as monomers, principally because they

appear in this form in crystal structures of ligand-receptor complexes. However, TNF-R1 has also been crystallized as both head-to-head and head-to-tail dimers [45], and there is genetic and experimental evidence that Fas, TNF-R1 and CD40 exist as preformed oligomers within the plasma membrane [46]. Self-association of the receptors depends on an N-terminal pre-ligand association domain (FLAD) that includes the first CRD and that is not directly involved in ligand binding.

TNFR-related structures - the EGF-like domain
A1 and B2 modules are not restricted to the TNF receptor family but also form the structural basis of epidermal growth factor (EGF)-like domains present in several proteins such as laminins. Laminins are composed of three related chains (α , β and γ , of which there are different isoforms) associated by a C-terminal coiled-coil domain. These chains display several globular domains in their N-terminal moieties with intervening, elongated structures composed of EGF-like repeats [47]. As shown in Fig. 4, the overall structure of EGF-like and CRD repeats is strikingly similar, except that A1-B2 modules in laminin are separated by an additional module, which we have designated S1 because of its small size [48]. EGF-like repeats 3-5 of the γ 1 chain of laminin bind Nidogen-1, a protein that interconnects laminin molecules in the basement membrane, but whose sequence is unrelated to TNF or C1q family members. So far, there is no evidence that EGF-like repeats bind TNF- or C1q-like ligands.